AU/ACSC/YAROVINSKIY/AY17

AIR COMMAND AND STAFF COLLEGE

AIR UNIVERSITY

INTELLIGENCE SURVEILLANCE AND RECONNAISSANCE FULL MOTION VIDEO AUTOMATIC ANOMALY DETECTION OF CROWD MOVEMENTS: SYSTEM REQUIREMENTS FOR AIRBORNE APPLICATION

by

Aleksandr Yarovinskiy, Maj, USAF

A Research Report Submitted to the Faculty

In Partial Fulfillment of the Graduation Requirements for the Degree of

MASTER OF MILITARY OPERATIONAL ART AND SCIENCE

Advisors: Dr. Andrew Niesiobedzki, Dr. Robert Smith

Maxwell Air Force Base, Alabama

October 2017

Disclaimer

The views expressed in this academic research paper are those of the author and do not reflect the official policy or position of the US government or the Department of Defense. In accordance with Air Force Instruction 51-303, it is not copyrighted, but is the property of the United States government.

Table of Contents

List of Tables

List of Figures

Abstract

The collection of Intelligence, Surveillance, and Reconnaissance (ISR) Full Motion Video (FMV) is growing at an exponential rate, and the manual processing of it cannot keep up with its growth. The purpose of this study is to develop automatic solutions to help analysts produce actionable intelligence for the warfighter. This paper will address the question of how can automatic pattern extraction, based on computer vision, extract anomalies in crowd behavior in ISR imagery. This paper will overview recent advances in automatic crowd anomaly detection techniques and the current technology necessary to implement them in the field. Assumptions are made for linear and ideal scaling of crowd anomaly detection techniques, using current technology, for field applications. The end product is a proposed pod system for airborne applications capable of processing an area the size of a small city for all crowd anomalies, and transmission of results to a ground node. Further study is required to optimize the proposed system for efficiency of scale.

Introduction

Overview of the Study

Automated processing of Intelligence, Surveillance and Reconnaissance (ISR) Full Motion Video (FMV) is necessary to keep up with an exponentially increasing collection of imagery and its manual processing. New automated methods are being published continuously to process FMV and they should be utilized to cue intelligence analysts to areas if interest. It is not the intention of this study to replace human talent, but rather utilize it more efficiently with emerging technology of crowd anomaly detection techniques. This study will assess this technology and propose a system capable of processing an area in the permissive environment to cue analysts to areas of interest.

Nature of the Problem

In 2001, 255 terabytes (TB) was transmitted per month, today it has grown to 1,300 TB, and the next generation of sensors will collect 2,200 TB per day.[1] The amount of data collected over the years shows exponential growth, and due to manual processing, requires exponential growth of manpower to keep up with the process. It is unsustainable to expand manual processing of exponentially increasing data, so the collected imagery remains un-used and unable to contribute to the warfighter needs.

The permissive environment, which the proposed system is geared for, allows freedom of movement for the general population. Most of the time the analysts are looking for a "needle in a haystack," the one insurgent (or a very small group of them) hiding among a crowd, making preparations for their next operation. The analyst first needs to figure out the general behavior of

the crowd, and then what doesn't fit in. That makes the process very tedious and time-consuming. Meanwhile, the imagery being collected is piling up and waiting to be processed.

Purpose of the Study

The purpose of the study is to help the analyst, mentioned in the previous section, using automation. The methods this paper proposes will help the analyst determine what doesn't fit in the general behavior of the crowd, taking into consideration time of day, predominant crowd patterns, weather, and other environmental factors. The analyst will then determine which anomalous events are of intelligence value if any.

Research Question

The question that this paper attempts to address is: based on computer vision, how can automatic pattern detection extract anomalies from crowd behavior in ISR FMV imagery? It is not the intent of the paper to extract intelligence via automatic means, just anomalies that may have intelligence value, as determined by actual analysts.

Definition of Terms

Algorithm: A set of equations that are programmed in a computer system to determine crowd anomalies. Has unique properties of speed and accuracy, within a given imagery sequence.

Algorithm Performance: The proposed algorithms will be compared for their speed and accuracy against the same dataset and standardized across the same hardware.[2] Algorithm speed is measured in seconds per frame (spf) or frames per second (fps), and algorithm accuracy is measured in percentage rating of how accurately it performs against truth data.

Computer Vision: Current software capabilities allow computers to recognize shapes in the video images. This is a general term used to describe this capability and where to take it to next.

Crowd Anomaly: Anything in the crowd that is different from predominant crowd behavior, as determined by statistics. For example, a person moving in the opposite direction of the main crowd, or at a speed much slower or much faster than the main crowd. A vehicle or an odd piece of equipment among the main crowd. Anything that a reasonable analyst may find odd.

GPU Node/Cluster: Graphics Processing Unit (GPU) computing. Within the last 10 years, graphics cards used extensively by computer gamers, proved orders of magnitude performance speed-up over traditional central processing unit (CPU) computing. A GPU node consists of a computing system comprising of 4-8 GPU processors. A cluster is a networked system of many nodes.

Permissive Environment: Area of operations which allows freedom of movement of friendly assets.

Resolution requirement: System resolution has a specific minimal threshold necessary for an algorithm to recognize a human shape.

Sensor array: Sensor is a photoelectric device with the capability to capture and pixilate an image. Array is a group of sensors arranged in a square or rectangle. A cell phone camera is an example of a sensor for the purpose of this paper.

Spatial Anomaly: Discriminant saliency of the occurrence, such as high or low speed, or opposite direction of a pedestrian within a crowd.

Stand-off Distance: The distance of the airborne platform for which this system is proposed, as measured from a sensor array to an area being observed.

Supervised/Unsupervised Learning: Algorithms can learn anomalies based on labeled sequences, derived from truth data, to establish their statistical model. Truth data needs to be parsed by a human analyst to teach an algorithm what is, and what is not an anomaly. Supervised learning algorithms make extensive use of truth data to establish the statistical model. The more truth data is processed, the more precise the algorithm will be. Unsupervised learning algorithms rely on comparison of observed data to previous sequences, thereby establishing adaptable learning ability.

Temporal Anomaly: An event of low probability. For example, a pedestrian walking alone may be considered temporal anomaly because it is a rare occurrence.

Truth data: That pattern that is defined by a human analyst as a true anomaly.

Research Methodology

This paper will present problem/solution framework to identify the problem of ISR FMV processing gap and offer automation solution to augment human analysts. The criteria for a solution are found in Strategic Master Plan (SMP) for the next 20 years. Analysis of alternatives will overview a series of published algorithms for anomaly detection and compare them using performance metrics. Best performing algorithms will be selected and implemented using existing hardware. The solution will be evaluated using scaled performance metrics and its capabilities against criteria from Strategic Master Plan. System recommendations will be presented on its optimization and field utility.

Background and Significance

Background

> "A visitor to a DCGS node might see an NCO sitting in front of a row of screens watching video. That airman "may be supporting a task, a 'pattern-of-life' development— it could be a lot of things," James said. "But if he's watching video ... I would offer that's a lousy use of the human brain.""[3]

The above excerpt illustrates the manual processing of ISR imagery. The comment was made by Lt Gen Larry James, Air Force ISR chief, suggesting that there is a better way to process imagery than looking at it manually. The process involves looking at FMV imagery for extended periods of time and extracting information out of it that would be of use to the warfighter.

Crowd anomaly detection is a new field in computer vision, and combined with better algorithm implementation over the years, shows promise for application in the field. This paper will go over various crowd anomaly detection algorithms published to date using primary scholarly sources to compare the performance data of various implementations. The sources which were selected for comparison used standardized imagery data[4], so that a fair comparison could be made. Based on comparison results detailed, a proposed system would be scaled for implementation in the field.

Significance

The premise behind extracting anomalies in crowds has to do with the fact that terrorists, insurgents, or bad actors, in general, like to hide within crowds.[5] However, they have a different agenda than the rest of the population and hence behave differently in terms of movement, congregation, or equipment on hand, thus contributing to anomalous patterns in crowds. For example, overlooking a given crowd over an extended period of time, an analyst can come up with crowd behavior patterns. The analyst can establish peak hours of movement during start and end of a work day, account for environmental factors, and establish predominant speed and

direction of traffic. An anomaly in a pattern could be a vehicle at odd hours of the day, suggesting abnormal activity. Another anomaly could be a person moving within a crowd in the wrong direction, or at a speed much faster or slower than the others. What can they be up to? That is something that is difficult to find manually but may hold some intelligence value.

Automatic anomaly extraction methods will allow analysts to concentrate on potential interest areas, rather than looking for them. Not all anomaly extraction results will be of intelligence value, and further manual processing will be necessary to distinguish anomalous pattern results.

Criteria for Success

The following section deals with criteria for success in problem/solution framework for the proposed system.

Strategic Master Plan Criteria

The Air Force Future Operating Concept (AFFOC) makes references for Strategic Master Plan capabilities necessary in the next 20 years. The following capability requirements, as referenced in SMP will be used as criteria for success in this research:

1. Price, in reference to ISR.1—"…focus on moderately priced systems, to include commercial technology, for permissive environment";
2. Capability, in reference to ISR.4—"…enhance capabilities to holistically detect, monitor, analyze, and attribute threats…and improve target systems analysis…";
3. Modularity and interoperability, in reference to AG2.1—modularity requirements.[6]

Resolution Criteria

There are two aspects of resolution—optical and digital. Optical resolution is governed by how big the aperture of a lens is, and digital resolution is the pixilation of an image into a digital form. If either resolution requirements do not meet the minimum threshold, automatic crowd anomaly extraction algorithm will not work. Since the proposed system will be airborne, it is necessary to size optics and pixilation for the operating altitude of the airborne platform, hence the stand-off distance.

Resolution requirements are defined for the optical system as a function of stand-off distance, which is necessary for the correct performance of anomaly detection in crowds. The goal of anomaly detection methods is to detect anomalous human activities in crowds. Hence optics have to be sized to resolve human shapes at desired stand-off distances.[7] Furthermore, area coverage will be addressed for the proposed system.[8]

Airborne Application Criteria

To further comply with SMP AG2.1, the proposed system will feature 2000lb class pod requirement, so that any airborne platform that can carry 2000lb class munition externally can carry this custom 2000lb class pod, which would house modular components necessary for system operation. Most of the system weight will comprise of the data processing units, called nodes, which will do the main processing of the imagery. The true resolution of the proposed system is vast, and it will not be feasible to transmit all of the imagery. Rather, the imagery will be down sampled on-board and transmitted along with high-resolution portions of it, comprising only of the anomalies.

Analysis of Alternatives

Overview

Anomaly detection in crowds is an emerging field, and better and faster algorithms are published continuously. This research will concentrate on the performance of anomaly detection methods when applied to a standardized dataset against standardized hardware.[9] The dataset consists of security camera footage fixed on a pedestrian walkway on University of California, San Diego (UCSD) campus. The performance of the algorithms is measured by how fast and how accurately they can pick out the anomalous activity within the crowd of pedestrians. Anomalous patterns can include but are not limited to cyclists, skateboarders, wheelchairs, vehicles, as well as people walking on grass. An example sequence is depicted in Figure 1, which highlights the crowd pattern anomalies at about 75% detection accuracy.[10] The idea is to use automation to pick out abnormal activity, something that is different from predominant behavior. When applied in the field, abnormal behavior may consist of loitering, reconnoitering, moving within a crowd erratically, congregating during abnormal hours and in abnormal locations, or carrying bulky objects, among many others.

Method Comparison

Crowd anomaly detection algorithm performance will be overviewed in this section. The performance of algorithms is a measure of the speed with which a particular algorithm can find anomalies in the given imagery, as well as its accuracy against the truth data, defined earlier. The algorithms being overviewed are applied against standardized imagery set of UCSD campus. The anomalies are tagged as truth data, and it is the measure of the algorithms how quickly and how accurately can they pick out truth data, as tagged by the dataset author.

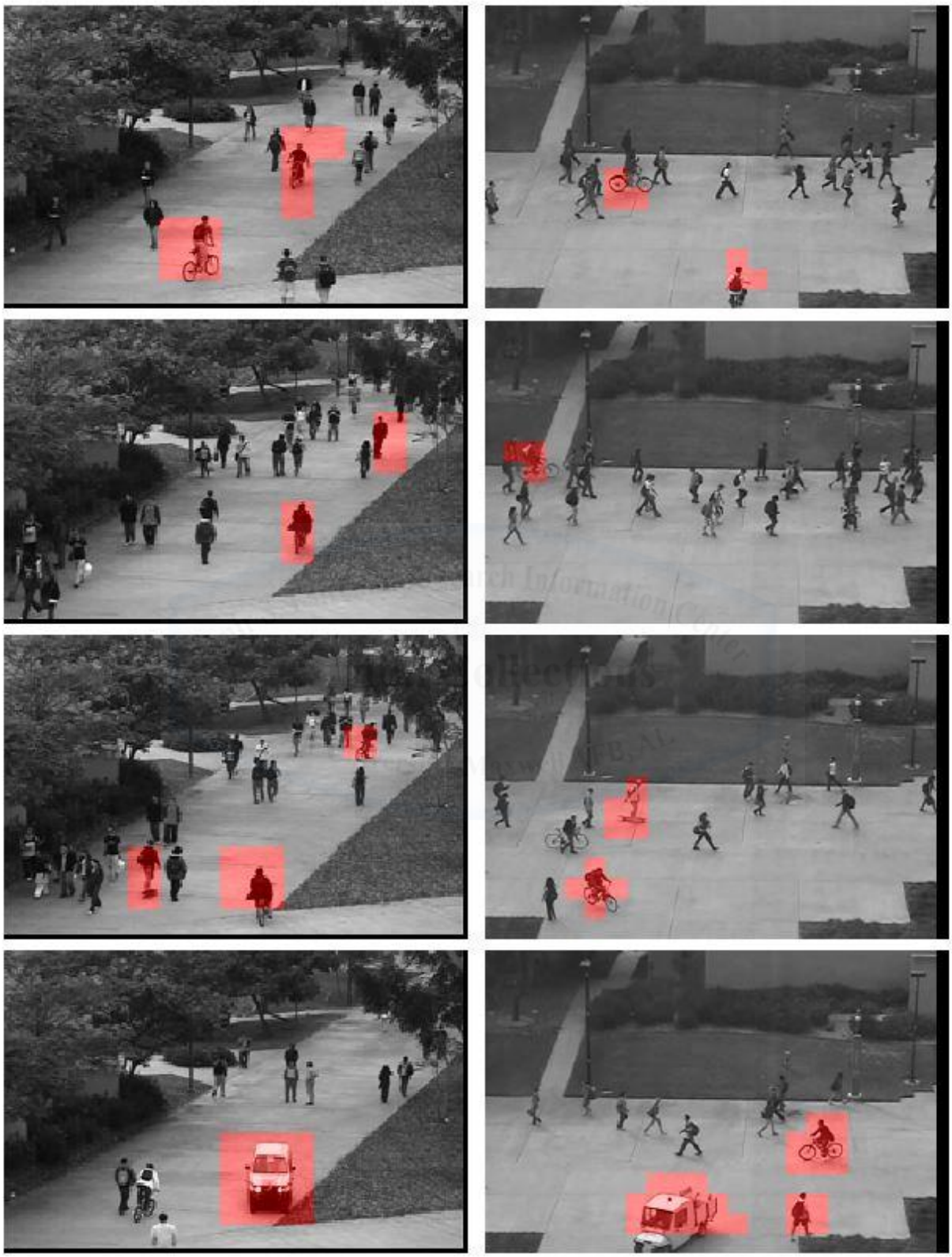

Figure 1: Anomaly Detection

Table 1: Anomaly detection methods comparison

	PED1 dataset		PED2 dataset		
Anomaly Detection Method[11]	Speed [spf]	Frame EER [%]	Speed [spf]	Frame EER [%]	Pixel EER [%]
Mahadevan, 2010[12]	21	25	29	25	55
Reddy, 2011[13]	0.07	22.5	0.097	20	32
Roshtkhari, 2013[14]	0.19	15	0.22	NA	29
Lu, 2013[15]	0.007	15	0.0096	NA	42
Li, 2014[16]	0.81	NA	1.01	18.5	29.9
Xiao, 2015[17]	0.28	NA	0.37	10	17
Gunduz, 2016[18]	0.023	29	0.011	15	NA
Sabokrou, 2016[19]	NA	NA	**0.0027**	11	**15**
Sabokrou, 2017[20]	0.074	NA	0.099	**8.2**	19

First, the algorithms are taught to pick out shapes that move with high or low differential speed in relation to other shapes, like bicycles which move faster than the crowd or loiterers that move slower. The algorithms also recognize which areas have very low probability of traffic, like grass. The implications of this technology are to pick irregular movement of persons, as compared to predominant crowd behavior, which may, or may not, hold intelligence value.

Table 1 shows comparison of various anomaly detection methods. The speed is given in seconds per frame (spf), and the lower number is better. Frame equal error rate (EER) is how accurately the algorithm can pick out frames that contain anomalies in the video sequence, and pixel EER is how accurately those anomalies can be localized within a frame. Both lower EER

ratings are better. The best performing method ratings are boldfaced. Also, PED1 and PED2 have different resolution, hence different performance rating.[21] Algorithm performance and accuracy have improved over the years.

The above algorithms can be grouped into two categories: supervised and unsupervised learning. Each has specific advantages associated with it, but can be likened to a novice versus an advanced performer. What this means is that when a novice learns his job, first he follows the basics the way it was taught in school. This is analogous to a supervised algorithm. It makes decisions based on preloaded rules set by analysts. Unsupervised algorithms make statistical models based on predominant behavior and can adjust their anomaly rules based on evolving imagery data. This is useful in the field when the insurgents change their tactics.

Supervised Algorithms

Mahadevan, 2010 algorithm is the first to incorporate detection of anomalies, of spatial and temporal nature, in the same technique. Its predecessors used staged imagery to detect only spatial or temporal anomalies. This method was the first to utilize a natural environment of UCSD campus walkway to look for both types of anomalies. In the included imagery in Figure 1, an example of a temporal anomaly is the vehicle in the pedestrian walkway on the bottom two images. It has a low probability of occurrence, per the definition above, and the field equivalent to that could be a group of insurgents during odd hours of the night conducting earth excavations to emplace an Improvised Explosive Device (IED) along the patrol route of friendly forces. Spatial anomalies in crowd movement are the events that exhibit statistically different properties than the predominant pattern. In the image sequence above, the cyclists and skateboarders are highlighted as anomalies because of their differential speed in relation to pedestrians. The direction of movement could also be a factor, but since in this particular dataset the pedestrians

move in either direction, it's not. Field equivalent application could be detection of an illegal transaction in a crowded marketplace.

Reddy, 2011 algorithm utilizes an advanced implementation of the same techniques found in Mahadevan, 2010. Its efficiency comes from dividing input frames into tiles, and processing them separately as if they are independent image sets. Once anomalies are found and identified, the results are stitched together. This type of implementation can be used on high resolution imagery, where each image can be divided into tiles, and each tile sent to a separate computing node, enabling the computer processing cluster to scale linearly.

Lu, 2013 method demonstrates unrivaled speed using MATLAB language software to extract anomalies using CPU-architecture computing. This method would show great promise if it were to be implemented on GPU-architecture computing. The proposed system has room to grow in terms of performance and scalability, given the evolving nature of newer algorithms.

Li, 2014 is an improved version of the method used in Mahadevan, 2010. It was written by the same researchers and shows the progress of the effort of anomaly detection in spatial and temporal domains. The improved version of the algorithm demonstrates improved accuracy of the anomaly detection method using filters based on where anomalies occur. This filtering technique demonstrated the accuracy increase by 50 percent. This source is presented for comparison purposes and does not hold significant edge over its competitors.

Gunduz, 2016 method has an interesting artifact: its speed improves with the higher resolution of the imagery. This can be a typo, or it is possible that it has an optimal performance rating at a specific resolution. Regardless of the fact, the method possesses an additional feature of crowd density awareness and seems to have a better error rate with the higher resolution

imagery dataset. Another interesting feature of this method is that its fps performance is on par with Lu, 2013. This improvement could be further translated into great potential if this algorithm were to work on GPU computing architecture.

Sabokrou, 2016 and 2017 are the winning methods in this research and will be utilized in the proposed system requirements. The 2016 version of this method was published to only work with GPU computing architecture. It showed orders of magnitude better performance than its predecessors. The 2017 paper made a side by side comparison of this algorithm working on both CPU and GPU architectures. Its fps performance on CPU architecture is not the fastest, as can be seen, when comparing Lu, 2013 and Gunduz, 2016, which implies there is significant room for improvement and optimization, should both of the mentioned methods were made to work on GPU architectures.

Unsupervised Algorithms

Roshtkhari, 2013 is the first published implementation of anomaly detection using unsupervised learning model. It relies on the observation of dominant behavior of crowds and can detect spatio-temporal anomalies dynamically based on dominant behavior. That means that the fielded system does not need to process hours of sample imagery to learn patterns, but can be deployed right away to learn predominant patterns of each subculture of the region. This implementation can be useful in the field where geography plays a major part on the tactics of the insurgents.

Xiao, 2015 method is an accuracy improvement to the Roshtkhari, 2013 algorithm. Both methods utilize unsupervised learning, but Xiao, 2015 demonstrates significant improvement in accuracy, with a frame error rate of 10% and localization error rate of an anomaly at 17%. Figure 2 illustrates this improvement in accuracy. The man walking the bicycle is detected in

this algorithm (Figure 2, lower left), but not in Reddy, 2011 (Figure 1, right col, third down). That is an example of frame EER performance, where an entire anomaly was detected within a frame (even thought that particular frame contained three anomalies). Figure 2 also illustrates how anomalies are better contained within the highlighted part of the image, whereas Figure 1 anomalies are highlighted coarsely. That is an example of pixel EER performance. The proposed system in the field will only transmit anomalies with high resolution within downsampled imagery, so pixel EER performance translates into telemetry bandwidth savings.

Figure 2: Improved Anomaly Detection

Scaled System Comparison

Table 2 illustrates the scaled performance of the proposed system.[22] The ratings for each method are in frames per second (fps). The higher number is better, and the best ratings are boldfaced for each system architecture. Performance is inversely proportional to the size of the optical pixel sensor array. The bigger the array, the greater the area coverage, but slower fps rating.[23] All the methods for anomaly detection are implemented on central processing unit (CPU) architecture. Sabokrou, 2016 algorithm is the first known method to be implemented in

both CPU and graphics processing unit (GPU), with significant demonstrated speed-up. The new technology trend has shown the emergence of GPU clusters, which offer significant computation speed-up to the previous generation of CPU-only computing.[24] What is also evident in this table is that Sabokrou 2017, the same algorithm which has been implemented on both GPU and CPU architectures, is not the fastest among its CPU-implemented peers. The future of anomaly detection may see significant performance improvements when other anomaly detection methods transition to GPU-accelerated computing. That said, the proposed system will utilize Sabokrou, 2016 method on a GPU-accelerated cluster.

Table 2: Scaled system performance comparison

Anomaly Detection Method	3 x HPc7000 96-processor 2.5GHz Intel Xeon-2679[25]			30 x Quantum TXR111-2000R 120 x Tesla P100 GPU Cluster[26]		
	Speed [fps]			Speed [fps]		
Sensor Array[27]	1x1	3x3	5x5	1x1	3x3	5x5
Reddy, 2011[28]	2.19	0.24	0.09	NA	NA	NA
Roshtkhari, 2013[29]	0.96	0.11	0.04	NA	NA	NA
Lu, 2013[30]	**21.97**	**2.44**	**0.88**	NA	NA	NA
Li, 2014[31]	0.21	0.02	0.01	NA	NA	NA
Xiao, 2015[32]	0.57	0.06	0.02	NA	NA	NA
Gunduz, 2016[33]	19.96	2.22	0.8	NA	NA	NA
Sabokrou, 2016[34]	NA	NA	NA	**213.3**	**23.7**	**8.53**
Sabokrou, 2017[35]	2.14	0.24	0.09	NA	NA	NA

Proposed System

Area Coverage and Performance

Table 3 illustrates the proposed anomaly detection system area coverage and performance. As the number of sensors are added for the increased area coverage, the speed rating of anomaly detection algorithm decreases for that area. It's an inverse proportional relationship.

Table 3: Area coverage and performance

Sensor Array	Length Coverage [m]	Width Coverage [m]	Speed [fps]
1x1	504	783	213.3
3x3	1512	2350	23.7
5x5	2520	3916	8.53
7x7	3528	5482	4.35
10x10	5040	7832	2.13

System Dimensions and Requirements

Table 4 illustrates the entire proposed system. The number of computing nodes were chosen for successful performance at 8.5fps of the specified anomaly detection method,[36] at desired area coverage, 2.5km x 4km. This area can cover most places of interest in asymmetric warfare operations to look for anomalies in crowds. Optics requirements for the proposed system are not covered, but requirements for the desired stand-off distances are sized in Table 5 to ensure proper operation of anomaly detection method in crowds. Its area coverage is limited by the resolution of the sensor array, illustrated in Table 3 and can be calculated by Equation 6 in Appendix D. The accuracy is not a real measure of performance of the proposed system in the field. The accuracy is derived from the published algorithm for a standardized dataset for UCSD pedestrian walkway. Actual accuracy of the system will depend on algorithm learning from

sample ISR imagery. The accuracy will also depend on the natural environment, as the proposed system is airborne and is subject to weather patterns.

Table 4: Overall proposed system specifications

Computing	30 x Quantum TXR111-2000R nodes utilizing 120 x Tesla P100 GPU processors
Effective system stand-off	See Table 5 for optics limitation
System area coverage	See Table 3 for area coverage and performance trade-offs
System accuracy	91.8% frame anomaly detection and 85% pixel localization[37]
System weight	30 nodes x 50lb each, 500lb overhead for optics, cables, etc. Overall 2000lb class custom pod
System size	30 computing nodes occupy 30U of 19" server rack space Overall 55"x22"x30" for computing nodes Additional optics dimensions not specified
System power	60kW for computing nodes, 40kW overhead Total 100kW power requirement
System cooling	Externally mounted 2000lb class slotted pod, natural cooling
System telemetry and recording	Self-contained commercial-of-the-shelf (COTS) items

The total weight is estimated by the weight of 30 computing nodes at 50lb each and is given a fudge factor to account for other equipment. Its size is estimated by the overall dimensions of the computing nodes. Optical sensors will occupy additional space and possibly weigh more. Total power requirement is derived from a full operating load of 30 computing nodes and additional overhead of 40kW for redundancy and other equipment. Since the computing nodes will need 60kW of power, all of that power will be converted to heat at full

load. The suggested cooling method is natural with slotted pod design. The pod will be mounted externally on an airborne platform and will be flying at a high altitude where the air is moving, and it's cold. The pod will also feature self-contained telemetry and recording equipment to send down-sampled imagery with highlighted pattern anomalies to the ground node or satellite uplink. It will not require integration to other avionics systems of its host platform.

System Optimization

The proposed system performance is limited by the implemented algorithms. As stated earlier, anomaly detection in crowds is an emerging field. Its methods are getting better every year. This system architecture is GPU-accelerated, using the only anomaly detection algorithm published to date with implemented GPU-accelerated computing. This same algorithm, when implemented on legacy architecture is not the fastest, and when faster algorithms[38] come online for GPU-accelerated systems, they can be implemented on a proposed system, so there is much room for improvement and optimization.

Current anomaly detection methods require labeled frame sequences to establish the internal statistical model. Newer algorithms will not need learning from labeled sequences and will learn on the go.[39] That means the more they will be used in the field, the better they will get at determining anomalous patterns. Analysts will always be needed to help algorithms learn better and to determine if anomalous patterns are of intelligence value, instead of just looking for "patterns-of-life" mentioned earlier.

The proposed system was scaled linearly with ideal efficiency. It is a matter of the fact that GPU-accelerated systems don't scale linearly, but in fact logarithmically.[40] Some system performance improvements may include subdividing frame images into tiles and sending each tile as separate video imagery for each node, a kind of "divide-and-conquer" approach. This

approach was demonstrated using Reddy, 2011 algorithm, which showed several orders of magnitude improvement. When nodes finish processing the tiles for anomalies, they can be stitched back together.

Unified Memory Networks is a system optimization technique to consolidate system memory into a common pool so that it is addressable by all processors.[41] GPU processors are optimized to work independently because they have not been designed for networked computing, but for individual hobbyists playing computer games. Recently, they have been adapted to be used for cluster computing, but share system limitations. It is similar to using a Formula 1 car on public roads. One of the proposed solutions for GPU cluster optimization is a virtual network for the memory, which comes on-board of each GPU unit.

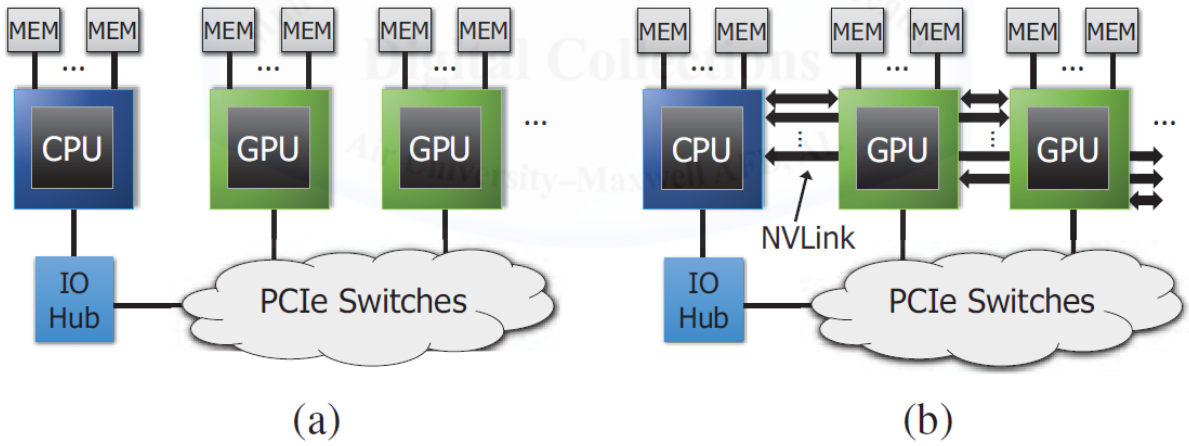

Figure 3: (a) Traditional GPU node architecture (b) Proposed Unified Memory Network NVLink by NVIDIA[42]

Figure 3 on the left shows traditional GPU system architecture. The GPU processor comes on a board with included memory. It interfaces with the node CPU through PCIe Switches and IO hub, which are the system limitation. It is like having a speed limit for the Formula 1 car on a public road, which in fact can go much faster. The portion on the right adds the NVLink, so that

each GPU can address the other's memory and share the load, bypassing the slow PCIe switches. The above approaches demonstrated system memory latency reduction rates in excess of 75% and PCI-e bus speed-up of 10 times, all contributing to the improvements in efficiency in GPU cluster scalability.

As mentioned earlier, system accuracy is limited, in part, by weather patterns. Specifically, at large stand-off distances, atmospheric distortion plays a role in image clarity. If images are not clear, they will contribute to poor system accuracy. One way to overcome system accuracy due to atmospheric distortion is to introduce additional image processing technique called super-resolution.[43] Super-resolution algorithms clear up noisy or blurry images, but they have not been accounted in the performance calculations of the proposed system due to poor standardization technique to compare them. Super-resolution method based on efficient sub-pixel convolution network (ESPCN) relies on redundant video imagery to clear up noise. Since FMV consists of many frames which have the same content, but from a little different angle, this method will utilize image variations to learn imagery gradients and produce crisper images real-time. Super-resolution will not extract additional detail out of imagery that cannot be captured due to resolution limitations, but if there is noise in the image due to atmospheric distortion, it will clear it up for analysts to review, or bump it up to the threshold necessary for the anomaly detection algorithms to work effectively. It is an option when system accuracy falls below the accepted threshold, but it will impact fps performance ratings.

The proposed system has a requirement of 100kW of power. Most airborne platforms do not provide this much power to auxiliary components. Fortunately, existing products are available on the market that can supply power generated by the airstream.[44] The COTS generator is in the form of an additional 500lb pod that can be mounted to an external weapon

station and will generate power from the oncoming airstream. The system will come online during stable flight, when maximum airstream and natural cooling are available. The only modification to existing airborne platforms is to run power cables from the power generating pod to the proposed system.

System Utilization Scenario: Drug Deal Detection

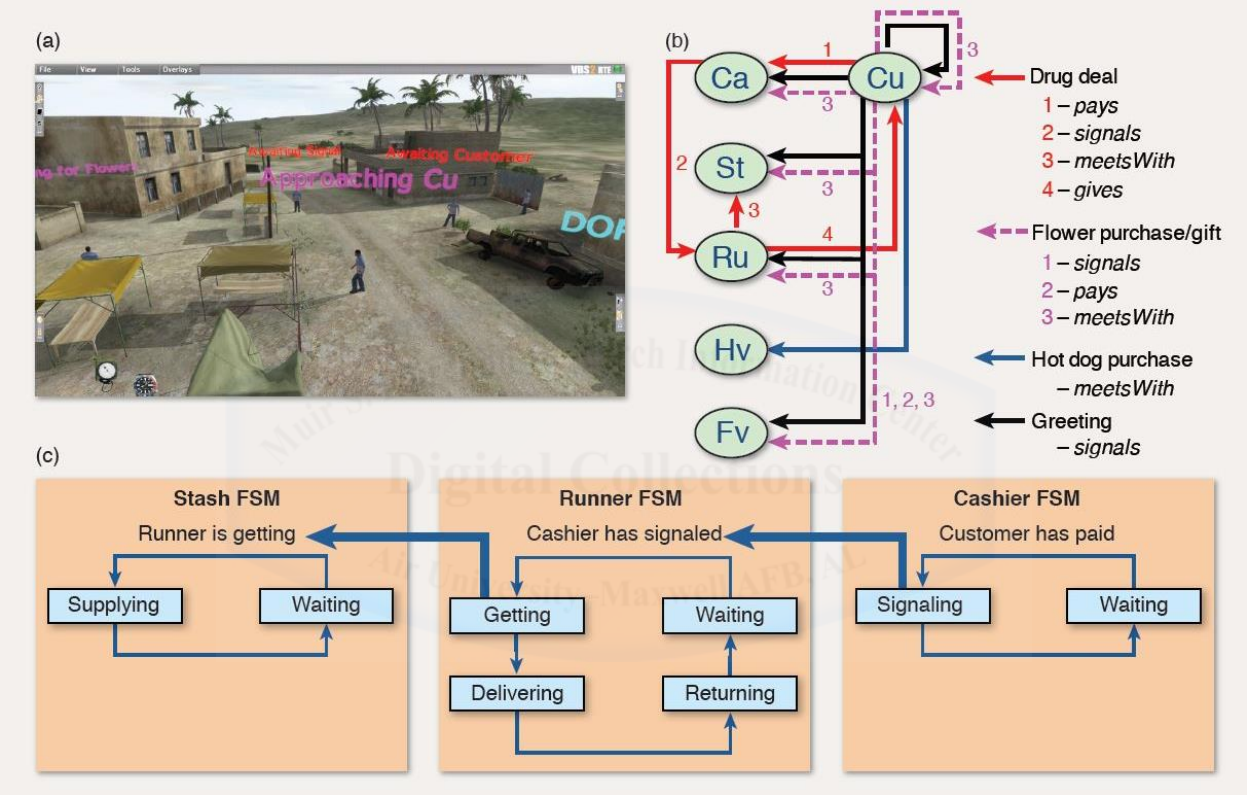

Figure 4: Drug Deal Transaction

(a) Drug Transaction Simulation in VBS2. (b) All parties involved, including bystanders and other vendors in a marketplace: customer (Cu), cashier (Ca), runner (Ru), stash (St), hotdog vendor (Hv), and flower (Fv). (c) Transaction requirements

The following scenario will examine the proposed system utility in detecting a drug transaction in a crowded marketplace. It has long been established that drug trade has supported other illegal activities in asymmetric warfare environment, in the form of the opium trade in Afghanistan, as an example.[45]

Figure 4(a) illustrates a drug transaction scenario that was generated in Virtual Battlespace 2 gaming environment.[46] There are several steps necessary to accomplish successful drug transaction for all parties so as not to tip off authorities. They tend to be complicated and involve several parties. In this case, the customer initiates a transaction with the cashier and makes the payment (1). Once the cashier is satisfied, he signals to the runner (2) to get the goods from the stash (3) and deliver to the customer (4). Meanwhile, other vendors are also in the area along with other customers, represented by hotdog and flower vendors. The flower purchases, greetings, and transactions are part of normal crowd behavior. It's the runner, who's getting the goods from the stash and supplying the customer that can be considered a spatial crowd anomaly. This can be quantified, highlighted and sent for analysis with the proposed system for anomaly detection of crowd patterns.

System Utilization Scenario: IED Emplacement Detection

Allied forces are patrolling an area known for insurgent activity. Convoys have encountered numerous IEDs on the main travel route. The area is close to a village, and the proposed system for anomaly detection of crowd patterns is deployed on an Unmanned Aerial Vehicle (UAV). This system has already downloaded sample imagery from other ISR platforms and learned the patterns of crowd behavior. It patrols the area persistently for 24 hours, and it knows what time people should be going to work, what the main travel routes are, and at what speed they should be traveling. It logs activity (a temporal anomaly) along the main route for Allied patrols. The time of activity is late at night or early in the morning, way before the predominant crowd pattern travels, and the activity includes several males congregating and conducting earth excavations. This system has the coverage of the large area, 2.5km x 4km, and it also saw which direction this group of people was heading from

and going to. It was abnormal for them to travel at this time of night and this activity was logged in as well. These anomalies were transmitted using built-in data-link to the intelligence processing node, and the analysts determined the activity was another IED emplacement. The ISR platform was re-routed to look where the group of individuals was headed. Night raid was scheduled for that location, and Explosive Ordnance Disposal (EOD) technicians were called to blow up IED in-place.

Conclusions

SMP Criteria

The SMP states ISR.1 criteria to "…focus on moderately priced systems, to include commercial technology, for permissive environment".[47] The proposed anomaly detection system is built entirely out of COTS items. The optical sensor is available on the market, and so are computing nodes. The custom components are optics system and pod design, relatively low-cost items. This system will also run on existing algorithms for crowd anomaly detection. There will be no need for research and development of new software techniques, only the adaptation of existing methods.

The SMP states ISR.4 criteria to "…enhance capabilities to detect, monitor, analyze, and attribute threats holistically…and improve target systems analysis…"[48] The proposed system contributes to this criteria. This will not solve all ISR imagery processing issues, but analysts will concentrate on crowd pattern anomalies, rather than looking for them. This will help to close the gap between imagery collection and imagery processing.

The SMP calls for systems modularity in AG2.1.[49] The proposed system is of modular design, its size and weight specified for 2000lb custom pod mounted externally. Most aircraft in

the Air Force inventory, which carry externally mounted 2000lb munitions, should handle this capability.

Resolution Criteria

The proposed system was designed for anomaly detection algorithms to work on an area of 2.5km by 4km. This is the area equivalent to a small city and can be processed by a system sizeable for 2000lb class pod at a rate of 8.5 frames per second. Stand-off requirement will call for optics of appropriate design, which are provided in Appendix C.

Recommendations

Automated processing of ISR FMV is necessary to keep up with the growing demand. The current technology makes it possible to collect data at an unprecedented rate, and its capabilities are growing exponentially. It is not possible to keep up with this growth, while still using manual means to process imagery. This problem has been addressed in this research by proposing automated means to process ISR FMV to cue analysts of possible areas of interest in the permissive environment. There are limitations to this system, due to its inherent need for specific resolution for anomaly extraction algorithms to work accurately. This system is designed to process imagery in a permissive environment. It has to be able to resolve human shapes, so it needs certain digital resolution capabilities as well as optical resolution.

The proposed system is not meant to look for targets, which is pattern detection. It is meant to analyze all crowds and determine what doesn't fit into the general pattern of crowd behavior. It is assumed that people move about their business in a predictable way in a village or a small city under surveillance. There are peak hours when people go to and from work; the crowds move in a predictable direction at a predictable speed. The crowd anomaly can be defined by a person who is not moving at that speed, makes frequent stops, or moves in the

wrong direction. This type of anomaly can be passed to an analyst who may discount it as noise or may dig into it to find clues for illicit operations. It is meant to help the analysts filter out the noise when dealing with insurgents, who hide among the civilian population.

Anomaly detection algorithms have been studied using a standard dataset in a natural environment, a walkway in UCSD campus. Within this environment, anomalies have been defined as cyclists, skateboarders, wheelchairs, vehicles on the sidewalk, and pedestrians walking on grass. This type of classification can be extended and superimposed on asymmetric warfare environment. We know that insurgents will not move with the main crowd to conduct their operations. They will prepare for operations, conduct meetings, use messengers and carry equipment to prepare for the next assault. Their purpose and their behavior will give them away if they try to hide within crowds.

There are two types of algorithms for anomaly detection, supervised and unsupervised. A supervised algorithm requires labeled imagery sequence as truth data to learn the difference between imagery with crowd anomalies and without them. This type of algorithm may be a good starting point for system implementation. There are thousands of hours of imagery already collected that can be labeled for algorithms to start learning, and then they can be used "out-of-the-box". Unsupervised algorithms rely on observing the natural environment and see what does not fit. They can be more adaptable towards the changing environment. For example, suppose one day all students in UCSD campus decided to ride bicycles on the walkway. The unsupervised algorithm would adapt to this environment, and now the pedestrians would be considered as crowd anomalies. This type of algorithm carries with it advanced implications; it can change and adapt to rapidly evolving battlefield.

The proposed system utilizes existing computer software algorithms and current hardware. No advanced research is necessary to build it. However, this does not guarantee its success. Optics and digital resolution have to be matched to specific spatial requirements to guarantee successful operation of anomaly detection algorithms. Extensive testing has to be completed to ensure the system operates as predicted with base lined anomalies in scaled-up capacity. Finally, operational techniques have to be developed for the field to define what the crowd anomalies in asymmetric warfare environment are.

The proposed system requirements are based on currently available performance information of algorithms and hardware. The paper showed a trend of continuous improvements of algorithm speed performance and accuracy. It also discussed future hardware optimization techniques to improve the efficiency of scalability. The 2000lb weight and 100kW power requirements are just the starting points on the road to miniaturization.

There are several layers of software for this system to be effective. Firmware is an overarching term to update system functionality and components. That will happen continuously as it does with most electronic systems. The system will also need to be updated continuously with truth data to make it more accurate. Also, when an algorithm is changed, it would have to go through learning and validation process, which will be in the form of block updates.

There is a compelling need for such a system in the operational environment to close the gap between imagery collected and imagery processed. When imagery is processed in an expedient way, it becomes actionable intelligence that can be used to save lives and win battles against insurgents before they ever start.

Appendix A

Focal Length

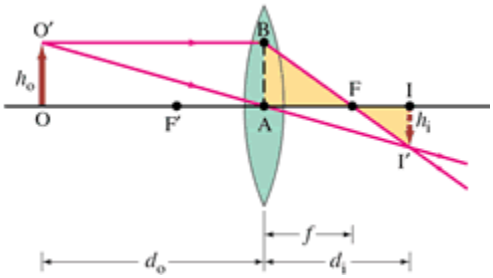

Figure 5: Convex lens ray tracing basics

Ray tracing is used to determine the position of an image when viewed through convex lens.[50] In this paper ray tracing will be used to determine focal length, *f*, as a function of standoff distance, d_o.

In Figure 5, the referenced parameters are listed below:

f = focal length of lens

II' = image on the optical sensor

OO' = real object observed

d_o = standoff distance

d_i = distance to optical sensor

h_o = required resolution for detection, 4cm=0.04m [51]

h_i = CMOS sensor distance between adjacent pixels, 0.0016mm=1.6E-6m [52]

From Figure 5, triangles O'AO and I'AI are similar, hence their dimensions are proportional. We can derive the following equation:

$$\frac{d_o}{d_i} = \frac{h_o}{h_i}$$

Equation 1

Since $d_o >> d_i$, focal length f can be approximated to d_i. Re-arranging the terms for focal length as a function of standoff distance will give us the following equation:

$$f = \frac{h_i d_o}{h_o}$$

Equation 2

Appendix B

Aperture

The Rayleigh criterion defines the minimum angle at which point objects can be resolved, due to diffraction limit of light, and is defined by the following equation[53]:

$$\theta = 1.22\frac{\lambda}{D}$$

Equation 3

where λ is the wavelength of light, D is the aperture diameter and θ is the minimum angle of resolution. This paper will use green light for calculation of aperture, given it is the midpoint of visible light spectrum with λ = 555nm. From Figure 5, angle O'AO = θ, and the standoff distance relationship is given below:

$$\theta = tan^{-1}\frac{h_o}{d_o}$$

Equation 4

The previous equations can be combined and solved for aperture size as a function of standoff distance:

$$D = 1.22\frac{\lambda}{tan^{-1}\frac{h_o}{d_o}}$$

Equation 5

Appendix C

Optics Calculation

Table 5: Standoff distance optics calculation

Standoff [m]	Focal length [mm]	Aperture [mm]
1000	40	17
2000	80	34
3000	120	51
4000	160	68
5000	200	85
6000	240	102
7000	280	118
8000	320	135
9000	360	152
10000	400	169
11000	440	186
12000	480	203
13000	520	220
14000	560	237
15000	600	254
16000	640	271
17000	680	288
18000	720	305
19000	760	322
20000	800	339

Appendix D

System Scaling

All algorithm times, t_i, presented in this research are scaled against Intel Core 2 Quad Q9550 CPU. The scaling method used is based on MDT calculation time presented in Roshtkhari paper, which uses specified hardware.[54] All research papers use MDT algorithm[55] as a baseline for comparison using their own unspecified hardware, hence its run time alone can predict the scaling comparison against known hardware, as well as scale factor for proposed method.

Table 6: Processor performance

Hardware	Performance score[56]	Tflops performance[57]
Intel Core 2 Quad Q9550 CPU	4011	NA
2.5 GHz Intel Xeon-2679 v4	25236	NA
Nvidia GeForce Titan X Pascal	NA	0.343
Nvidia Tesla P100	NA	4.7

Table 6 is presented to calculate processor performance factor, *PPF*, necessary for system scaling. It is the ratio of performance of processor used in the proposed system to the processor performance of the source literature. In this table, PPF_{CPU} = 25236/4011= 6.3, PPF_{GPU} = 4.7/0.343= 13.7.

Table 7: Resolution

Resolution [pixels]	Horizontal	Vertical	Total
PED1	240	160	38,400
PED2	360	240	86,400
250Mp sensor	19580	12600	246,708,000

Table 7 presents resolution that will be used for scaling purposes in the proposed system. It will be designated as R_{PED1}, R_{PED2}, R_{250Mp}, in total pixels format.[58]

Area coverage [meters dimension], will be given by the number, *n*, of 250Mp sensors and its number of pixels in length, *l*, and width, *w*, direction, multiplied by 4cm linear distance resolution requirement:

$$l = 12600 * 0.04 * n_{sensors}$$
$$w = 19580 * 0.04 * n_{sensors}$$

Equation 6

Finally, scaled performance of presented algorithms in the proposed system, in FPS_i (frames per second), is given in the following equation:

$$FPS_i = \frac{R_{PED_i} n_{processors}}{t_i R_{250Mp} n_{sensors}} PPF * SF$$

Equation 7

Where *SF* is efficiency scaling factor, assumed to be 1 (ideal linear scaling)

Notes

[1] (James 2012, 8)
[2] Appendix D details standardization method
[3] (Schanz 2013, 24)
[4] (UCSD 2013)
[5] (Joint Publication (JP) 3-0 2011, V-12)
[6] (Secretary of the US Air Force 2015, 39-44)
[7] Appendices A-C detail optics size, for resolution requirements specified, as a function of desired standoff distances
[8] Appendix D covers formulas for calculating area coverage
[9] (UCSD 2013), Appendix D Table 7
[10] (Reddy, Sanderson and Lovell 2011, 59)
[11] Performance speed is scaled per Appendix D
[12] (Mahadevan, et al. 2010)
[13] (Reddy, Sanderson and Lovell 2011)
[14] (Roshtkhari and Levine 2013)
[15] (Lu, Shi and Jia 2013)
[16] (Li, Mahadevan and Vasconcelos 2014)
[17] (Xiao, Zhang and Zha 2015)
[18] (Gunduz, et al. 2016)
[19] (Sabokrou, Fayyaz, et al., Deep-Anomaly: Fully Convolutional Neural Network for Fast Anomaly Detection in Crowded Scenes 2016)
[20] (Sabokrou, Fayyaz, et al., Deep-Cascade: Cascading 3D Deep Neural Networks for Fast Anomaly Detection and Localization in Crowded Scenes 2017)
[21] See Appendix D Table 7 for PED1 and PED2 resolution
[22] Scaled performance is calculated per Appendix D Equation 7
[23] Appendix D Equation 6 and Equation 7
[24] (NVidia 2017, 40)
[25] (Hewlett Packard Enterprise 2017)
[26] (Exxact Corp. n.d.)
[27] (Canon 2015)
[28] (Reddy, Sanderson and Lovell 2011)
[29] (Roshtkhari and Levine 2013)
[30] (Lu, Shi and Jia 2013)
[31] (Li, Mahadevan and Vasconcelos 2014)
[32] (Xiao, Zhang and Zha 2015)
[33] (Gunduz, et al. 2016)
[34] (Sabokrou, Fayyaz, et al., Deep-Anomaly: Fully Convolutional Neural Network for Fast Anomaly Detection in Crowded Scenes 2016)
[35] (Sabokrou, Fayyaz, et al., Deep-Cascade: Cascading 3D Deep Neural Networks for Fast Anomaly Detection and Localization in Crowded Scenes 2017)
[36] (Sabokrou, Fayyaz, et al., Deep-Anomaly: Fully Convolutional Neural Network for Fast Anomaly Detection in Crowded Scenes 2016)
[37] (Sabokrou, Fayyaz, et al., Deep-Cascade: Cascading 3D Deep Neural Networks for Fast Anomaly Detection and Localization in Crowded Scenes 2017), (Sabokrou, Fayyaz, et al., Deep-Anomaly: Fully Convolutional Neural Network for Fast Anomaly Detection in Crowded Scenes 2016)
[38] (Lu, Shi and Jia 2013), (Gunduz, et al. 2016)
[39] (Roshtkhari and Levine 2013), (Xiao, Zhang and Zha 2015)
[40] (Beri, Bansal and Kumar 2017),
[41] (Zhan, et al. 2016), (Kim, et al. 2014)
[42] (Kim, et al. 2014, 485)
[43] (Caballero, et al. 2017)
[44] (ATGI n.d.)
[45] (US Joint Forces Command 2010, 61)
[46] (Lin, et al. 2011, 50)

[47] (Secretary of the US Air Force 2015, 42)
[48] (Secretary of the US Air Force 2015, 40)
[49] (Secretary of the US Air Force 2015, 42)
[50] (Pedrotti n.d.)
[51] (Axis Communications 2013)
[52] Sensor dimensions 20.2x29.2mm, sensor resolution 12600x19580, adjacent pixel distance is calculated by dividing dimensions by resolution and taking the worst case, which is vertical (Canon 2015)
[53] (OpenStax College 2013)
[54] (Roshtkhari and Levine 2013)
[55] (Mahadevan, et al. 2010)
[56] (PassMark Software n.d.)
[57] (Microway n.d.)
[58] (UCSD 2013)

Bibliography

ATGI. "ATGI Hi-power Ram Air Turbine." n.d. http://atgi.us/products-and-services/airborne-pod-systems/ (accessed May 2017).

Axis Communications. "Perfect Pixel Count: Meeting your operational requirements." May 2013. https://www.axis.com/files/feature_articles/ar_perfect_pixel_count_55971_en_1402_lo.pdf (accessed May 2017).

Beri, Tarun, Sorav Bansal, and Subodh Kumar. "The Unicorn Runtime: Efficient Distributed Shared Memory Programming for Hybrid CPU-GPU Clusters." *IEEE Transactions on Parallel and Distributed Systems*, May 2017: 1518-1534.

Caballero, Jose, et al. "Real-Time Video Super-Resolution with Spatio-Temporal Networks and Motion Compensation." Apr 2017. https://arxiv.org/abs/1611.05250 (accessed May 2017).

Canon. "Canon develops APS-H-size CMOS sensor with approximately 250 megapixels." Sep 7, 2015. http://global.canon/en/news/2015/sep07e.html (accessed May 2017).

Exxact Corp. *Tesla GPU Cluster Quantum TXR113-1000R specifications.* n.d. https://exxactcorp.com/NVIDIA-Tesla/GPU-Cluster.php (accessed May 2017).

Gunduz, Ayse Elvan, Cihan Ongun, Tugba Taskaya Temizel, and Alptekin Temizel. "Density aware anomaly detection in crowded scenes." *IET Computer Vision*, Aug 2016: 374-381.

Hewlett Packard Enterprise. "HPE BladeSystem c7000 Enclosure." Apr 2017. https://www.hpe.com/h20195/v2/GetPDF.aspx/c04128339.pdf (accessed May 2017).

—. "HPE ProLiant BL460c Gen9 Server Blade." Mar 2017. https://www.hpe.com/h20195/v2/GetPDF.aspx/c04380273.pdf (accessed May 2017).

James, Larry D., Lt Gen. "Airmen: Delivering Decision Advantage." *Air & Space Power Journal*, Nov-Dec 2012: 4-11.

Joint Publication (JP) 3-0. *Joint Operations.* Aug 11, 2011.

Kim, Gwangsun, Minseok Lee, Jiyun Jeong, and John Kim. "Multi-GPU System Design with Memory Networks." *MICRO*. Cambridge: IEEE, 2014. 484-495.

Li, Weixin, Vijay Mahadevan, and Nuno Vasconcelos. "Anomaly Detection and Localization in Crowded Scenes." *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Jan 2014: 18-32.

Lin, Jeffrey S, Ariel M Greenberg, Clayton R Fink, Wayne L Bethea, and Andrea G Savvides. "Group Activity Analysis for Persistent Surveillance." *Johns Hopkins APL Technical Digest*, 2011: 47-57.

Lu, Cewu, Jianping Shi, and Jiaya Jia. "Abnormal Event Detection at 150 FPS in MATLAB." *ICCV*. Sydney: IEEE, 2013. 2720-2727.

Mahadevan, Vijay, Weixin Li, Viral Bhalodia, and Nuno Vasconcelos. "Anomaly detection in crowded scenes." *CVPR*. San Francisco: IEEE, 2010. 1975-1981.

Microway. *Comparison of NVIDIA Tesla/Quadro and NVIDIA GeForce GPUs.* n.d. https://www.microway.com/knowledge-center-articles/comparison-of-nvidia-geforce-gpus-and-nvidia-tesla-gpus/ (accessed May 2017).

NVidia. "NVidia Testla P100 The Most Advanced Datacenter Accelerator Ever Built." 2017. https://images.nvidia.com/content/pdf/tesla/whitepaper/pascal-architecture-whitepaper.pdf (accessed May 2017).

OpenStax College. "Limits of Resolution: The Rayleigh Criterion." Sep 2013. http://cnx.org/content/m42517/1.5/ (accessed May 2017).

PassMark Software. *CPU Benchmarks database.* n.d. https://www.cpubenchmark.net/cpu.php?cpu=Intel+Pentium+4+3.00GHz (accessed May 2017).

Pedrotti, Leno S. "Fundamentals of Photonics." n.d. https://spie.org/Documents/Publications/00%20STEP%20Module%2003.pdf (accessed May 2017).

Reddy, Vikas, Conrad Sanderson, and Brian C Lovell. "Improved anomaly detection in crowded scenes via cell-based analysis of foreground speed, size and texture." *CVPRW*. Colorado Springs: IEEE, 2011. 55-61.

Roshtkhari, M. Javan, and M. D. Levine. "An on-line, real-time learning method for detecting anomalies in videos." *Computer Vision and Image Understanding*, 2013: 1-17.

Sabokrou, Mohammad, Mohsen Fayyaz, Mahmood Fathy, and Reinhard Klette. "Deep-Anomaly: Fully Convolutional Neural Network for Fast Anomaly Detection in Crowded Scenes." Sep 2016. https://arxiv.org/pdf/1609.00866 (accessed May 2017).

—. "Deep-Cascade: Cascading 3D Deep Neural Networks for Fast Anomaly Detection and Localization in Crowded Scenes." *IEEE Transactions on Image Processing* , Apr 2017: 1992-2004.

Schanz, Marc. "ISR After Afghanistan." *Air Force Magazine*, Jan 2013: 22-27.

Secretary of the US Air Force. *Air Force Future Operations: A View of the Air Force in 2035.* Washington DC, Sep 2015.

UCSD. *UCSD Anomaly Detection Dataset.* Feb 2013. http://www.svcl.ucsd.edu/projects/anomaly/dataset.html (accessed May 2017).

US Joint Forces Command. *The Joint Operating Environment 2010.* Feb 18, 2010.

Xiao, Tan, Chao Zhang, and Hongbin Zha. "Learning to Detect Anomalies in Surveillance Video." *IEEE Signal Processing Letters*, Sep 2015: 1477-1481.

Zhan, Jia, et al. "A unified memory network architecture for in-memory computing in commodity servers." *MICRO*. Taipei: IEEE, 2016.